

Cyber infrastructure for *Fusarium*: three integrated platforms supporting strain identification, phylogenetics, comparative genomics and knowledge sharing

Bongsoo Park^{1,2}, Jongsun Park^{3,4}, Kyeong-Chae Cheong^{3,4}, Jaeyoung Choi^{3,4},
Kyongyong Jung^{3,4}, Donghan Kim^{3,4}, Yong-Hwan Lee^{3,4}, Todd J. Ward⁵,
Kerry O'Donnell^{5,*}, David M. Geiser^{1,2,*} and Seogchan Kang^{1,2,*}

¹Department of Plant Pathology, ²Integrative Biosciences Graduate Program in Bioinformatics and Genomics, The Pennsylvania State University, University Park, PA 16802, USA, ³Fungal Bioinformatics Laboratory, ⁴Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea and ⁵United States Department of Agriculture-Agricultural Research Service, Bacterial Foodborne Pathogens and Mycology Research Unit, Peoria, IL 61604, USA

Received August 15, 2010; Revised October 21, 2010; Accepted October 31, 2010

ABSTRACT

The fungal genus *Fusarium* includes many plant and/or animal pathogenic species and produces diverse toxins. Although accurate species identification is critical for managing such threats, it is difficult to identify *Fusarium* morphologically. Fortunately, extensive molecular phylogenetic studies, founded on well-preserved culture collections, have established a robust foundation for *Fusarium* classification. Genomes of four *Fusarium* species have been published with more being currently sequenced. The Cyber infrastructure for *Fusarium* (CiF; <http://www.fusariumdb.org/>) was built to support archiving and utilization of rapidly increasing data and knowledge and consists of Fusarium-ID, Fusarium Comparative Genomics Platform (FCGP) and Fusarium Community Platform (FCP). The Fusarium-ID archives phylogenetic marker sequences from most known species along with information associated with characterized isolates and supports strain identification and phylogenetic analyses. The FCGP currently archives five genomes from four species. Besides supporting genome browsing and

analysis, the FCGP presents computed characteristics of multiple gene families and functional groups. The Cart/Favorite function allows users to collect sequences from Fusarium-ID and the FCGP and analyze them later using multiple tools without requiring repeated copying-and-pasting of sequences. The FCP is designed to serve as an online community forum for sharing and preserving accumulated experience and knowledge to support future research and education.

INTRODUCTION

The fungal genus *Fusarium* poses a multifaceted threat to global crop production and animal/human health. Collectively, the genus includes many important plant pathogens (1). Certain *Fusarium* secondary metabolites, such as fumonisins, trichothecenes, enniatins and zearalenone, are toxins that threaten food safety and animal/human health (2). Some species infect immune-compromised individuals (3,4) but also cause corneal infections in people with healthy immune systems (5,6). Due to its practical importance, the genus has been extensively studied at levels ranging from genetic mechanisms underlying important traits, such as toxin production

*To whom correspondence (concerning phylogenetic data) should be addressed. Tel: +1 309 681 6383; Fax: +1 309 681 6672; Email: kerry.odonnell@ars.usda.gov

Correspondence (concerning phylogenetic data) may also be addressed to David M. Geiser. Tel: +1 814 865 9773; Fax: +1 814 863 7217; Email: dgeiser@psu.edu

Correspondence (concerning genomics data and informatics tools) may also be addressed to Seogchan Kang. Tel: +1 814 863 3846; Fax: +1 814 863 7217; Email: sxk55@psu.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

and pathogenicity, to global biodiversity and evolution (2,7–11).

More than 35 000 strains isolated from various substrates around the world are accessioned in the Fusarium Research Center (FRC) and the USDA-ARS National Center for Agricultural Utilization Research (NCAUR) Culture Collection, making this genus the best-preserved fungal group. Using this rich strain resource, extensive molecular phylogenetic studies have been conducted, resulting in data covering most agriculturally and/or medically important species complexes (6,9,10,12–22). However, despite these advances, a significant amount of diversity has yet to be explored, and some species complexes are quite poorly characterized phylogenetically. To support and coordinate the remaining phylogenetic analyses, it is essential to archive available phylogenetic data and associated cultures in a format that is readily accessible and searchable by members of the global *Fusarium* research community.

In 2004, we released Fusarium-ID, a simple, web-accessible BLAST server that consisted of sequences of the translation elongation factor 1 α (*EF-1 α*) gene from a fairly representative spectrum of *Fusarium* species (13). Since then, we have expanded Fusarium-ID to include sequences of multiple marker loci that represent almost all known species and to provide more data analysis and visualization tools. We also developed two additional

platforms, including the Fusarium Comparative Genomics Platform (FCGP) and the Fusarium Community Platform (FCP), to build a more comprehensive community resource, named as the Cyber infrastructure for *Fusarium* (CiF; <http://www.fusariumdb.org/>; Figure 1). The main motivation for building CiF was to support the archiving and integration of data and knowledge from disparate yet related areas of research on *Fusarium* through a single-integrated platform. Many informatics platforms supporting fungal research have been developed. However, because they are often specialized for only a subset of data (e.g. genome sequences, data associated with culture collections or specific gene families), integrated analysis of disparate data sets across multiple taxa is cumbersome, as data from multiple sources need to be mined and integrated in an *ad hoc* manner. With diverse, systematically archived data sets, the CiF aims to efficiently leverage new knowledge and support problem solving. We have been building platforms similar to CiF to support research and education on other taxa [e.g. the *Phytophthora* Database (23)]. These platforms share common architecture and tools, such that building new platforms and improving existing platforms is efficient and cost-effective. As more platforms are added, they will form a comprehensive cyber infrastructure supporting fungal research.

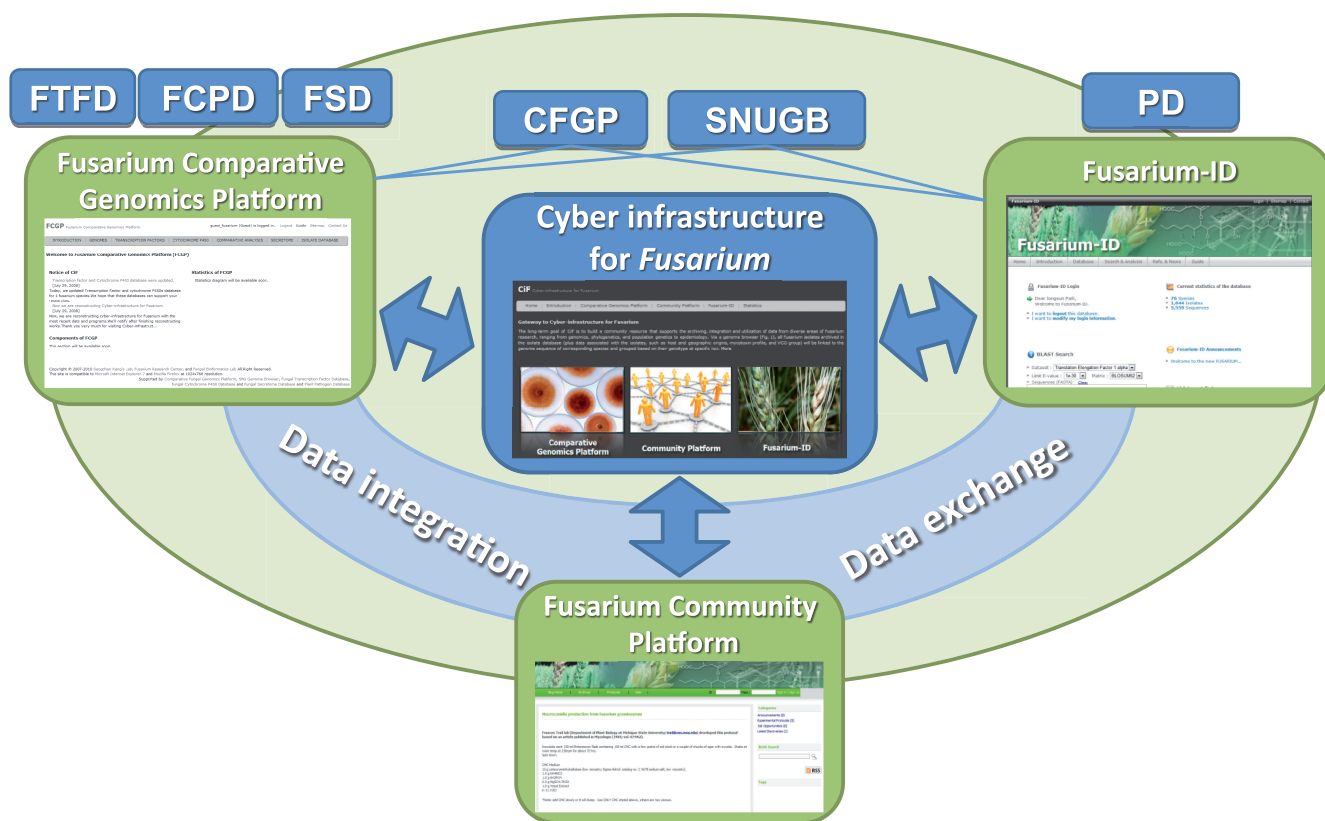


Figure 1. System structure of the CiF. The CiF consists of three components, FCGP, Fusarium-ID and FCP. Several databases specialized for fungal genome sequences or gene families and functional groups, including Fungal Transcription Factor Database (FTFD), Fungal Cytochrome P450 Database (FCPD), Fungal Secretome Database (FSD) and Comparative Fungal Genomics Platform (CFGP), feed data to the FCGP. Data analysis and visualization tools from the CFGP, the Seoul National University Genome Browser (SNUGB) and *Phytophthora* Database (PD) support data visualization and analysis.

FUSARIUM-ID

With the advent of molecular tools and robust molecular evolutionary principles, it has become easier and faster to recognize new species when they are encountered (24). Systematically archiving available *Fusarium* phylogenetic data will help guide future species descriptions, coordinate community research on its systematics, and support education in *Fusarium* biology. Without a robust phylogenetic framework and community-wide knowledge sharing, discovery and characterization of novel *Fusarium* species will likely be fragmented, creating confusion instead of the order that taxonomy should provide. The Fusarium-ID (<http://isolate.fusariumdb.org/>) consists of a database of extensive sequence data from most known *Fusarium* species and data analysis and visualization tools. The Fusarium-ID enables users to explore the diversity of *Fusarium* and accurately identify new isolates based on their sequence similarity to previously characterized species.

Data content and utility

The Fusarium-ID currently archives 5558 marker sequences from 1844 isolates representing over 200 phylogenetically distinct species. Its data content will grow rapidly, as we continuously curate and deposit data from previous and current phylogenetic studies. All sequence data in this database have been derived from vouchered, publicly available cultures, allowing users to further investigate any connections between their query and hits in the database. Most of the data in the Fusarium-ID database are also available and searchable through the Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Center (<http://www.cbs.knaw.nl/fusarium/>) so that they will be maintained in multiple electronic resources around the world.

Sequence data from more than 10 marker loci are archived in the Fusarium-ID database with some loci sequenced at multiple locations (e.g. ribosomal RNA encoding genes and their spacer regions), but individual species typically have been characterized only by a subset of these markers. However, sequences at three loci, including *EF-1 α* and two genes encoding the largest and second largest subunits of RNA polymerase (*RPB1* and *RPB2*, respectively), are being generated for all of the phylogenetic species, and can serve as markers individually and collectively to help identify new isolates to the species level. To identify a new isolate, sequence data from one or more of these genes can be used as a BLAST query against the Fusarium-ID sequence database. Considering the extensive coverage of human pathogenic species in the database, an exact match can be reasonably interpreted as definite species-level identification (10). However, because the value of a sequence match result depends on a few experimental or biological factors, all results must be interpreted with care (10,24). Precise conclusions may require phylogenetic analysis based on multiple markers, especially when the BLAST results suggest that the query sequence may represent a novel species not currently represented in Fusarium-ID. In such cases, users should employ an appropriate multilocus typing scheme to assess

genealogical concordance and evidence of genetic exchange (25,26). Individual sequences and sequence alignments from previous phylogenetic analyses can be downloaded from Fusarium-ID so that users can conduct their own analyses, and appropriate cultures can be ordered from FRC and/or the USDA-ARS NCAUR Culture Collection.

Data analysis and visualization tools

In addition to BLAST, Fusarium-ID provides a number of functions adopted from the *Phytophthora* Database (<http://www.phytophthoradb.org/>), a platform we developed to support the identification and monitoring of *Phytophthora* species and populations (23). Via the Folder function, users can create two types of data storage space in the CiF: (i) private folder for storing selected data and results from previous analyses and (ii) shared folder that permits data sharing with other users designated by the creator of the folder (by assigning user IDs permitted to access the folder). The shared folder function enables communication and collaboration among multiple users via the CiF. The Cart function described below allows users to collect data in multiple areas of Fusarium-ID and use/analyze them later. A suite of web tools, named the Phyloviewer, allows users to align sequences in BLAST outputs, including the query sequence, and any data stored in the Cart and build phylogenetic trees on the fly. Sequence data in the resulting tree are linked to information associated with corresponding isolates so that users can browse if any notable patterns (e.g. geographic and host of origins, mycotoxin profiles, etc) exist among the isolates included in the tree. The Virtual Gel function supports this diagnostic method by generating predicted restriction fragment length polymorphism (RFLP) patterns from chosen sequences and restriction enzyme(s) via a virtual gel.

Future plans

The CiF data warehouse will be continuously populated with phylogenetic sequence data from our previous and current studies to provide a robust foundation for ecological and phylogenetic studies and genome sequencing efforts. The utility of Fusarium-ID will be enhanced as members of the global *Fusarium* research community deposit cultures of novel species, along with associated sequence chromatograms and data, so that the sequence results can be verified and isolates are made available for future study. We also plan to add photographs and/or line drawings illustrating key morphological features associated with each phylogenetic species.

Two main functions that will be available soon are a geographic information system (GIS) tool and a tool for searching and comparing population genetic diversity data based on simple sequence repeat (SSR) loci. Both tools are currently functional in the *Phytophthora* Database. The GIS tool will function as a digitized atlas showing the genotypic and phenotypic diversity of *Fusarium* worldwide in geospatial and temporal contexts. This functionality will help to establish a baseline for monitoring the biogeographic diversity of *Fusarium* species. For major pathogenic *Fusarium* species, we plan to generate MLST

(Multi-Locus Sequence Typing) data sets for population-level analyses. In combination with the GIS tool and a search tool for MLST data (to be developed), users can significantly increase their sampling by integrating their datasets with those available in Fusarium-ID, monitor haplotype diversity across hosts and geographic regions, and examine the demographic history of species/populations.

FCGP

Rapidly accumulating genome sequence data from diverse *Fusarium* species with different traits offers tremendous opportunities for understanding the molecular and evolutionary mechanisms underpinning functional diversification at a genome level (8,27,28). The FCGP (<http://genomics.fusariumdb.org/>) was developed to facilitate the realization of such opportunities. Besides providing an interactive genome browser, the FCGP presents computed characteristics of multiple gene families and functional groups in sequenced species to support quick comparison and analysis across species. In combination with the phylogenetic framework and accessioned cultures available through Fusarium-ID, the FCGP will help users study the evolution of *Fusarium* genes, gene networks and whole genomes.

Genome data

The genomes of four *Fusarium* species, including *F. graminearum* (two strains), *F. oxysporum*, *F. verticillioides* and one species in the *F. solani* species complex, have been sequenced (8,27,28) with more species and isolates currently being sequenced or annotated. The first three species were sequenced by the Broad Institute (http://www.broadinstitute.org/annotation/genome/fusarium_group/MultiHome.html), while the Department of Energy Joint Genome Institute sequenced *F. solani* (also known as *Nectria haematococca* Mating Population VI; <http://genome.jgi-psf.org/Necha2/Necha2.home.html>). We converted sequences from these sites into a common format via the data extraction pipeline of the Comparative Fungal Genomics Platform (CFGP; <http://cftp.snu.ac.kr/>) (29). The most recent versions of genome data, as well as earlier versions, are available in the FCGP (under the 'List of all *Fusarium* genomes' in the 'GENOMES' menu). Results of the annotation of each gene with 12 different programs, including InterPro scans, subcellular localization prediction softwares, signal-peptide prediction programs and transmembrane helix prediction programs, are presented.

Genome browser

As the scale and complexity of genome sequence analyses increases, a versatile genome browser has become an essential tool. To support visualization and utilization of genome sequences and features both within and across species, the SNU Genome Browser (SNUGB; <http://genomebrowser.snu.ac.kr/>) (30), a genome browser developed to support the CFGP and several platforms derived from the CFGP, was integrated into the FCGP.

The application interface of the SNUGB was designed in a modular fashion to facilitate the addition of new tools and its customization for specialized platforms. The SNUGB has already been implemented in multiple platforms (31–37). All sequence data and contig information are displayed through the interface of the Contig Browser. Annotation information in a chosen region, such as transcripts, ORFs, tRNAs/rRNAs, exon/intron structure, SignalP, PSort and InterPro domains, can be displayed in multiple formats. Such information will be useful for in-depth analysis of gene function. In addition, the Chromosome Viewer shows the chromosomal locations of the phylogenetic markers stored in Fusarium-ID. Once the user clicks the bar indicating a specific marker, genomic sequences and features around the marker can be viewed through the SNUGB interface. The SNUGB can display genome sequences from multiple strains/species around a chosen locus to facilitate the evaluation and development of new phylogenetic markers, including the design of PCR primers.

Platforms that support the analysis of specific gene families and functional groups

We have constructed several comparative genomics platforms specialized for supporting in-depth analysis of specific gene families and functional groups in fungi (31,35,36). One of them is the Fungal Transcription Factor Database (FTFD; <http://ftfd.snu.ac.kr/>) (36), in which all putative transcription factors (TFs) encoded by sequenced fungal and oomycete species were identified and classified into families. Cytochrome P450s, a superfamily of heme-containing monooxygenases, play critical roles in fungal metabolism and ecology by participating in the production of diverse metabolites and also modifying harmful environmental chemicals (38). The Fungal Cytochrome P450 Database (FCPD; <http://p450.riceblast.snu.ac.kr/>) (35) archives genes encoding P450s to support studies on their function and evolution. The Fungal Secretome Database (FSD; <http://fsd.snu.ac.kr/>) identifies and archives putative secretory proteins (31). *Fusarium*-specific data from these three platforms, including 3075 TFs, 577 cytochrome P450s and 11668 putative secretory proteins, are organized to provide an overview of these proteins within and across *Fusarium* species. A BLAST server for each data set is available for quick search. Moreover, genes that appear unique to each species, as well as those that are present in subsets of the four species, were identified through BLASTMatrix2, a modified BLAST program that searches gene(s) homologous to a query in multiple species simultaneously. With FTFD, FCPD, FSD and BLASTMatrix2, *Fusarium* proteins can be quickly compared with those in other fungal taxa.

Future plans

In addition to depositing newly released *Fusarium* genome sequences, characteristics of additional protein groups, such as ABC transporters and carbohydrate degrading enzymes, will be added once the corresponding fungal kingdom-wide databases are established. Available

expressed sequence tags from *Fusarium* species will also be archived and linked to the corresponding genomes.

CART/FAVORITE: A HUB FOR DATA REPOSITORY AND ANALYSIS

Availability of multiple disparate, yet complementary, data through a single platform opens up the possibility of integrated analysis without going through data retrieval from multiple independent sources. However, to enable such analysis, a tool for retrieving and managing data from multiple databases is required. The Cart (or 'Favorite' in the FCGP) function serves as such a tool (Figure 2). Through this function, users can collect metadata for any sequences deposited in Fusarium-ID and the FCPD, and all or part of the collected data can be analyzed later by available data analysis tools, including BLAST, ClustalW, Virtual Gel, and Primer3, without requiring repeated copying-and-pasting of sequences for different analyses. Each cart/favorite can be stored in user's private and/or shared folder. The Cart (Favorite) function has been implemented in several platforms we have

developed (29,31,35,36). Many additional analysis tools that are currently operational in the CFGP (29) will be integrated to the CiF Cart/Favorite function to expand its utility.

FCP

Sharing experience and knowledge is fundamental to help leverage new knowledge and educate the next generation of researchers and educators. Without an efficient mechanism to support such effort, scientific endeavors will be fragmented and become inefficient. Web 2.0 technologies offer immense potential for supporting the pooling and sharing of diverse experience and knowledge accumulated in a global research community without being limited by distance and traditional forms of organizational structure. A number of scientific communities have built a web platform and associated databases and material resources to support community research and education. Some notable examples include the Arabidopsis Information Resource (TAIR; <http://www.arabidopsis.org/>), the iPlant Collaborative (iPlant; <http://www.iplantcollaborative.org/>), and the Plant Ontology Consortium (POC; <http://www.plantontology.org/>).

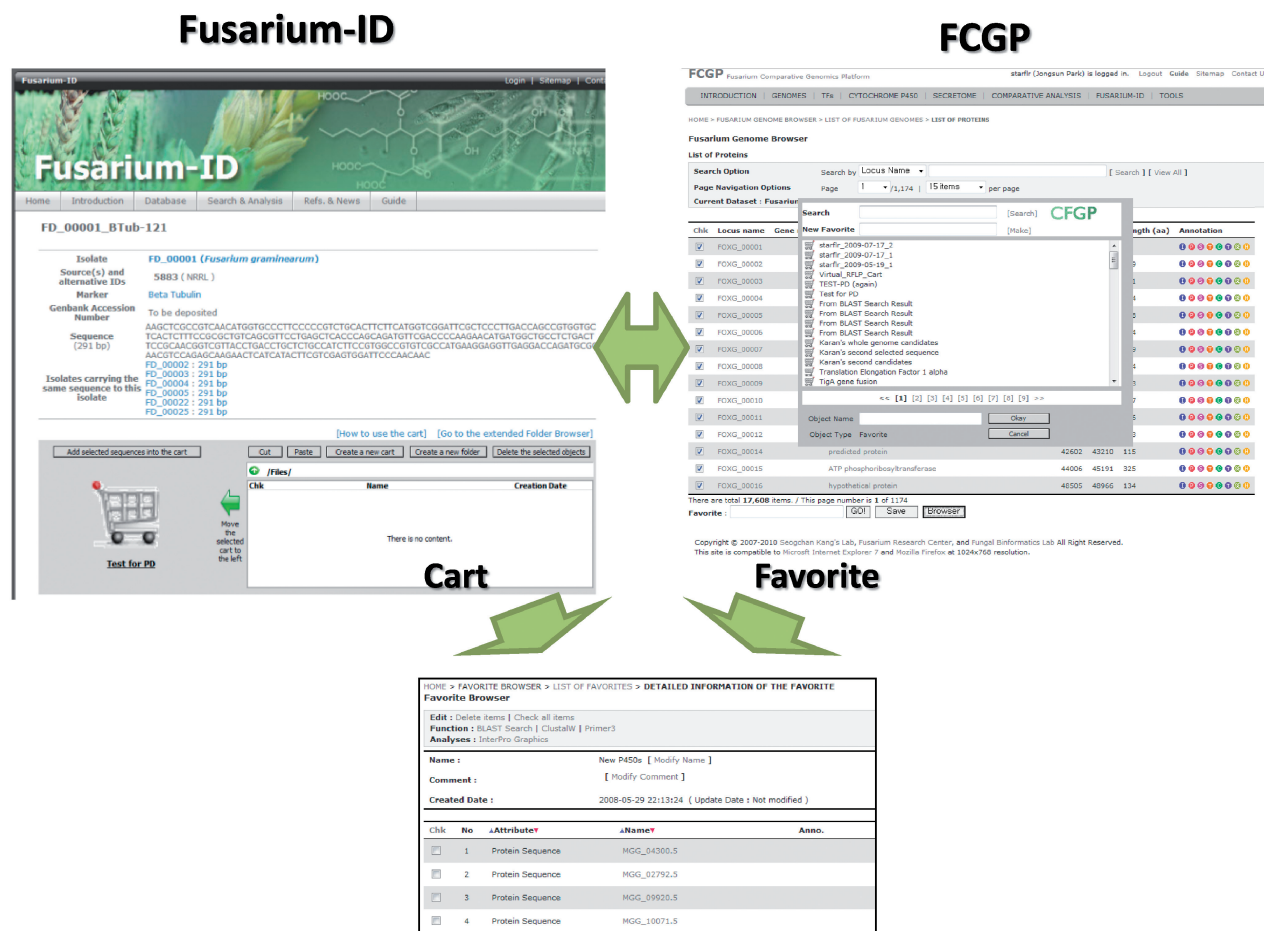


Figure 2. Cart/Favorite, a hub for data storage and analysis in the CiF. Users can collect sequences through two different interfaces: (i) Cart in Fusarium-ID and (ii) Favorite in the FCGP. All or part of the sequences collected in a cart/favorite, can be analyzed using several bioinformatics tools and can also be shared with other users through the Folder function.

www.iplantcollaborative.org/), the PHI-base (<http://www.phi-base.org/about.php>), GMOD (http://gmod.org/wiki/Main_Page) and the Aspergillus website (<http://www.aspergillus.org.uk/>). However, this potential needs to be more widely harnessed.

The FCP (<http://blog.fusariumdb.org/>) is currently being developed to support the preservation and sharing of experience and knowledge accumulated in the global Fusarium research community. Because the FCP is a new addition to the CiF, the currently available content is rather limited. However, through 'crowd sourcing' in the community, we plan to expand its content quickly. The FCP is expected to provide quick reviews of latest research development, experimental protocols, educational modules, and community news via a blog interface. The trail of communications associated with the archived content, particularly protocols, will help new comers to the community quickly learn from the collective knowledge rather than learning them through trial and error. Because it is web based, the FCP will also become an ideal medium for rapidly sharing information on emerging disease problems and coordinating subsequent responses. We envision that the FCP will also function to support global human networking.

AVAILABILITY AND REQUIREMENTS

All three platforms of the CiF can be accessed via the gateway web page (<http://www.fusariumdb.org/>). Although registration is not mandatory, certain functions, such as Folder and Cart and writing a post through the FCP, are not available without login. More than 250 users from 33 countries have registered. However, considering that guests can access most functions and data, the actual number of users and countries should be larger. User access data (monthly and yearly) can be found in the statistics page of the CiF.

ACKNOWLEDGEMENTS

We would like to thank Stacy Sink and Jean H. Juba for excellent technical assistance, Nathane Orwig for generating the DNA sequences in National Center for Agricultural Utilization Research (NCAUR)'s DNA core facility and individuals who have deposited *Fusarium* isolates to FRC and/or USDA-ARS-NCAUR.

FUNDING

This work was indirectly supported by grants from the United States Department of Agriculture-AFRI Plant Biosecurity program (grant numbers 2005-35605-15393 and 2008-55605-18773). These grants enabled the development of the *Phytophthora* Database. Tools and experience from the *Phytophthora* Database greatly facilitated the construction of the CiF. This work also was partially supported by the following grants to Y.-H.L.; National Research Foundation of Korea (grant numbers 2009-0063340 and 2009-0080161); Biogreen21 (grant number 20080401-034-044-009-01-00); Technology

Development Program for Agriculture and Forestry (grant number 309015-04-SB020); Crop Functional Genomics Center (grant number 2009K001198). Funding for open access charge: Penn State University.

Conflict of interest statement. None declared.

REFERENCES

- Agrios, G.N. (2005) *Plant Pathology*, 5 edn. Academic Press, San Diego.
- Desjardins, A.E. (2006) *Fusarium Mycotoxins: Chemistry, Genetics & Biology*. APS Press, St. Paul, MN.
- Anaissie, E., Kantarjian, H., Ro, J., Hopfer, R., Rolston, K., Fainstein, V. and Bodey, G. (1988) The emerging role of *Fusarium* infections in patients with cancer. *Medicine*, **67**, 77–83.
- Dignani, M.C. and Anaissie, E. (2004) Human fusariosis. *Clin. Microbiol. Infect.*, **10**(Suppl. 1), 67–75.
- Chang, D.C., Grant, G.B., O'Donnell, K., Wannemuehler, K.A., Noble-Wang, J., Rao, C.Y., Jacobson, L.M., Crowell, C.S., Sneed, R.S., Lewis, F.M.T. *et al.* (2006) Multistate outbreak of *Fusarium* keratitis associated with use of a contact lens solution. *JAMA*, **296**, 953–963.
- O'Donnell, K., Sarver, B., Brandt, M., Chang, D.C., Noble-Wang, J., Park, B.J., Sutton, D., Benjamin, L., Lindsley, M., Padhye, A. *et al.* (2007) Phylogenetic diversity and microsphere array-based genotyping of human pathogenic fusaria, including isolates from the multistate contact lens-associated U.S. keratitis outbreaks of 2005 and 2006. *J. Clin. Microbiol.*, **45**, 2235–2248.
- Balmas, V., Migheli, Q., Scherm, B., Garau, P., O'Donnell, K., Ceccherelli, G., Kang, S. and Geiser, D.M. (2010) Multilocus phylogenetics show high levels of endemic fusaria inhabiting Sardinian soils (Tyrrhenian Islands). *Mycologia*, **102**, 803–812.
- Ma, L.-J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B. *et al.* (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, **464**, 367–373.
- O'Donnell, K., Ward, T.J., Geiser, D.M., Kistler, H.C. and Aoki, T. (2004) Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. *Fungal Genet. Biol.*, **41**, 600–623.
- O'Donnell, K., Sutton, D.A., Rinaldi, M.G., Sarver, B.A.J., Balajee, S.A., Schroers, H.-J., Summerbell, R.C., Robert, V.A.R.G., Crous, P.W., Zhang, N. *et al.* (2010) An internet-accessible DNA sequence database for identifying fusaria from human and animal infections. *J. Clin. Microbiol.*, **48**, 3708–3718.
- Summerell, B.A., Leslie, J.F., Backhouse, D., Bryden, W.L. and Burgess, L.W. (eds). (2001) *Fusarium: Paul E. Nelson Memorial Symposium*. APS Press, St. Paul, MN.
- Aoki, T., O'Donnell, K. and Scandiani, M.M. (2005) Sudden death syndrome of soybean in South America is caused by four species of *Fusarium*: *Fusarium brasiliense* sp. nov., *F. cuneirostrum* sp. nov., *F. tucumaniae* and *F. virguliforme*. *Mycoscience*, **46**, 162–183.
- Geiser, D.M., Jiménez-Gasco, M., Kang, S., Makalowska, I., Veeraraghavan, N., Ward, T.J., Zhang, N., Kuldau, G.A. and O'Donnell, K. (2004) FUSARIUM-ID v.1.0: a DNA sequence database for identifying *Fusarium*. *Eur. J. Plant Pathol.*, **110**, 473–479.
- O'Donnell, K., Cigelnik, E. and Nirenberg, H.I. (1998) Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia*, **90**, 465–493.
- O'Donnell, K., Kistler, H.C., Cigelnik, E. and Ploetz, R.C. (1998) Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proc. Natl Acad. Sci. USA*, **95**, 2044–2049.

16. O'Donnell, K., Kistler, H.C., Tacke, B.K. and Casper, H.H. (2000) Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. *Proc. Natl Acad. Sci. USA*, **97**, 7905–7910.
17. O'Donnell, K., Nirenberg, H.I., Aoki, T. and Cigelnik, E. (2000) A multigene phylogeny of the *Gibberella fujikuroi* species complex: detection of additional phylogenetically distinct species. *Mycoscience*, **41**, 61–78.
18. O'Donnell, K., Sutton, D.A., Rinaldi, M.G., Magnon, K.C., Cox, P.A., Revankar, S.G., Sanche, S., Geiser, D.M., Juba, J.H., van Burik, J.-A.H. *et al.* (2004) Genetic diversity of human pathogenic members of the *Fusarium oxysporum* complex inferred from multilocus DNA sequence data and amplified fragment length polymorphism analyses: evidence for the recent dispersion of a geographically widespread clonal lineage and nosocomial origin. *J. Clin. Microbiol.*, **42**, 5109–5120.
19. Starkey, D.E., Ward, T.J., Aoki, T., Gale, L.R., Kistler, H.C., Geiser, D.M., Suga, H., Tóth, B., Varga, J. and O'Donnell, K. (2007) Global molecular surveillance reveals novel *Fusarium* head blight species and trichothecene toxin diversity. *Fungal Genet. Biol.*, **44**, 1191–1204.
20. Zhang, N., O'Donnell, K., Sutton, D.A., Nalim, F.A., Summerbell, R.C., Padhye, A.A. and Geiser, D.M. (2006) Members of the *Fusarium solani* species complex that cause infections in both humans and plants are common in the environment. *J. Clin. Microbiol.*, **44**, 2186–2190.
21. Schroers, H.-J., O'Donnell, K., Lamprecht, S.C., Kammeyer, P.L., Johnson, S., Sutton, D.A., Rinaldi, M.G., Geiser, D.M. and Summerbell, R.C. (2009) Taxonomy and phylogeny of the *Fusarium dimerum* species group. *Mycologia*, **101**, 44–70.
22. O'Donnell, K., Gueidan, C., Sink, S., Johnston, P.R., Crous, P.W., Glenn, A., Riley, R., Zitomer, N.C., Colyer, P., Waalwijk, C. *et al.* (2009) A two-locus DNA sequence database for typing plant and human pathogens within the *Fusarium oxysporum* species complex. *Fungal Genet. Biol.*, **46**, 936–948.
23. Park, J., Park, B., Veeraghavan, N., Jung, K., Lee, Y.-H., Blair, J.E., Geiser, D.M., Isard, S., Mansfield, M.A., Nikolaeva, E. *et al.* (2008) *Phytophthora* database: a forensic database supporting the identification and monitoring of *Phytophthora*. *Plant Disease*, **92**, 966–972.
24. Kang, S., Mansfield, M.A., Park, B., Geiser, D.M., Ivors, K.L., Coffey, M.D., Grunwald, N., Martin, F.M., Levesque, A. and Blair, J.E. (2010) The promise and pitfalls of sequence-based identification of plant pathogenic fungi and oomycetes. *Phytopathology*, **100**, 732–737.
25. Geiser, D.M. (2004) In Tkacz, J. and Lange, L. (eds), *Advances in Fungal Biotechnology for Industry*. Kluwer Academic Publishers, Dordrecht, pp. 1–12.
26. Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, D.S. and Fisher, M.C. (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genet. Biol.*, **31**, 21–32.
27. Coleman, J.J., Rounsley, S.D., Rodriguez-Carres, M., Kuo, A., Wasmann, C.C., Grimwood, J., Schmutz, J., Taga, M., White, G.J., Zhou, S. *et al.* (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet.*, **5**, e1000618.
28. Cuomo, C.A., Guldener, U., Xu, J.-R., Trail, F., Turgeon, B.G., Di Pietro, A., Walton, J.D., Ma, L.-J., Baker, S.E., Rep, M. *et al.* (2007) The *Fusarium graminearum* Genome Reveals a Link Between Localized Polymorphism and Pathogen Specialization. *Science*, **317**, 1400–1402.
29. Park, J., Park, B., Jung, K., Jang, S., Yu, K., Choi, J., Kong, S., Park, J., Kim, S., Kim, H. *et al.* (2008) CFGP: a web-based, comparative fungal genomics platform. *Nucl. Acids Res.*, **36**, D562–571.
30. Jung, K., Park, J., Choi, J., Park, B., Ahn, K., Choi, J., Kim, S., Choi, D., Kang, S. and Lee, Y.-H. (2008) SNUGB: a versatile genome browser supporting comparative and functional fungal genomics. *BMC Bioinformatics*, **9**, 586.
31. Choi, J., Park, J., Kim, D., Jung, K., Kang, S. and Lee, Y.-H. (2010) Fungal Secretome Database: integrated platform for annotation of fungal secretomes. *BMC Genomics*, **11**, 105.
32. Choi, J., Park, J., Jeon, J., Chi, M.H., Goh, J., Yoo, S.Y., Park, J., Jung, K., Kim, H., Park, S.Y. *et al.* (2007) Genome-wide analysis of T-DNA integration into the chromosomes of *Magnaporthe oryzae*. *Mol. Microbiol.*, **66**, 371–382.
33. Jeon, J., Park, S.Y., Chi, M.H., Choi, J., Park, J., Rho, H.S., Kim, S., Goh, J., Yoo, S., Choi, J. *et al.* (2007) Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. *Nat. Genet.*, **39**, 561–565.
34. Lee, W., Park, J., Choi, J., Jung, K., Park, B., Kim, D., Lee, J., Ahn, K., Song, W., Kang, S. *et al.* (2009) IMGd: an integrated platform supporting comparative genomics and phylogenetics of insect mitochondrial genomes. *BMC Genomics*, **10**, 148.
35. Park, J., Choi, J., Ahn, K., Park, B., Park, J., Kang, S. and Lee, Y.-H. (2008) Fungal cytochrome P450 database. *BMC Bioinformatics*, **9**, 402.
36. Park, J., Park, J., Jang, S., Kim, S., Kong, S., Choi, J., Ahn, K., Kim, J., Lee, S., Kim, S. *et al.* (2008) FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics*, **24**, 1024–1025.
37. Xi, H., Park, J., Ding, G., Lee, Y.-H. and Li, Y. (2009) SysPIMP: the web-based systematical platform for identifying human disease-related mutated sequences from mass spectrometry. *Nucl. Acids Res.*, **37**, D913–D920.
38. van den Brink, H.M., van Gorcom, R.F.M., van den Hondel, C.A.M.J.J. and Punt, P.J. (1998) Cytochrome P450 enzyme systems in fungi. *Fungal Genet. Biol.*, **23**, 1–17.